Orbital-optimized opposite-spin scaled second order correlation: An economical method to improve the description of open-shell molecules.

Rohini C. Lochan and Martin Head-Gordon

Department of Chemistry, University of California, and,

Chemical Sciences Division, Lawrence Berkeley National Laboratory,

Berkeley, CA 94720, USA.

Abstract.

Coupled cluster methods based on Brueckner orbitals are well-known to resolve the problems of symmetry-breaking and spin-contamination that are often associated with Hartree-Fock orbitals. However their computational cost is large enough to prevent application to large molecules. Here we present a simple approximation where the orbitals are optimized with the mean-field energy plus a correlation energy taken as the opposite-spin component of the second order many-body correlation energy, scaled by an empirically chosen parameter (recommended as 1.2 for general applications). This "optimized 2nd order opposite spin" (abbreviated as O2) method requires fourth order computation on each orbital iteration. O2 is shown to yield predictions of structure and frequencies for closed shell molecules that are very similar to scaled second order Moller-Plesset methods. However it yields substantial improvements for open shell molecules, where problems with spin-contamination and symmetry breaking are shown to be greatly reduced.

I. Introduction

Most standard wavefunction-based electron correlation treatments like second-order Mϕller-Plesset perturbation theor[1] (MP2) and coupled-cluster methods (CCSD[2], CCSD(T)[3]) typically employ restricted (RHF) or unrestricted (UHF) Hartree-Fock orbitals as their initial reference for closed- or open-shell systems, respectively. In general, the performance of these methods for closed-shell systems with respect to energetics, structural features and other molecular properties like vibrational frequencies are well documented and known to be quite reliable and accurate.[4] Typically, the description improves as the level of correlation included increases systematically (in the order MP2 < CCSD < CCSD(T)).[5] However, the reliability and systematic improvement of these methods that are based on UHF orbitals for open-shell systems is quite erratic.[6]

In particular, UMP2 (MP2 based on UHF orbitals) results are quite poor for predicting basic properties and reactivities of radicals.[6-9] This failure is largely attributed to the symmetry breaking problems or the spin-contamination of the UHF wavefunction, which is *not* an eigenfuction of the spin operator, $\hat{S}^2$.[10] For example, the $\langle S^2 \rangle$ values of the stable phenalenyl radical (a doublet) and sigma-complexed dimer (a singlet) have been evaluated to be about 2.1 and 3.2, respectively, at the UHF level of theory.[11] It is also known that the MPn series converges very slowly for cases where the underlying UHF wavefunction is highly spin-contaminated.[12-14]

In some processes like bond dissociation or when there is a small HOMO-LUMO gap associated with the system, the wavefunction may possess genuine multireference character. In such cases, a single determinantal reference is simply incapable of

describing the wavefunction, as opposed to the dissociation limit where unrestricted orbitals are a satisfactory single-reference description. We are then required to employ a multiconfigurational reference wavefunction like multiconfigurational self consistent field method (MC-SCF). However, the choice of configurations to be included, the need for additional corrections to include dynamic correlation and the related increase in computational complexity limit the use of such methods to small molecular systems.[15,16] Clearly, the quality of the reference wavefunction has a significant role to play in determining the performance of a method.

To tackle the spin contamination problems of the UHF orbitals one can switch to a restricted open shell HF (ROHF) wavefunction, which is an eigenfunction of $\hat{S}^2$ and thus eliminates spin-contamination in the reference determinant altogether. While ROHF does not permit easy description of the dissociation limit in hemolytic bond-breaking, restricted open shell MP2 (ROMP2)[17,18] and CC methods[19] have been implemented. These methods are reported to perform better than the corresponding unrestricted cases in describing structure and frequencies of radicals, especially for the cases where the spin contamination is high.[7,8] However, a statistical study on small doublet radicals indicates that ROMP2 continues to perform poorly. The authors attribute this failure of MP2 to HF orbital instabilities and symmetry breaking problems.[6]

Spin projection methods are also sometimes used to project out the leading spin contaminants from the converged (unrestricted) HF wavefunction[20,21] or at the MP level.[22] Alternatively reduction of spin-contamination and symmetry breaking problems due to the UHF wavefunction can be reduced by explicit inclusion of electron correlation. It is known that the deviation of the $\langle S^2 \rangle$ value from the actual limit of $S(S+1)$

decreases when the level of correlation is improved from UHF to UMP2 and UCCSD [23]. However, the decrease in case of UMP2 is not very significant.

A more drastic (and fundamental) solution is to eliminate the use of mean-field Hartree-Fock orbitals altogether. The use of optimized orbitals in coupled-cluster methods such as Brueckner coupled-cluster doubles (BD)[24,25] and orbital-optimized coupled-cluster doubles (OD)[26] were found to be effective in decreasing the spin contamination effects of the wavefunction and also in improving its stability towards symmetry breaking. However, these methods scale with the sixth power of system size and can be routinely applied to only relatively small molecules.

Kohn-Sham density functional theory (KS-DFT)[27,28] can also provide orbitals that are less prone to spin contamination as they include dynamic electron correlation effects during the orbital optimization procedure.[29] The KS-DFT description of high-spin open shell molecules is found to be rather good[6,30] and it can also be used as the reference wavefunction for coupled-cluster methods to obtain improved results at almost no additional cost.[31] However, present-day density functionals suffer from self-interaction errors[32] and also lack the ability to describe long-range van der Waals effects. Thus they cannot be used to study radical systems where either or both of these issues are important.[33] This failure of DFT is vividly portrayed by the recently studied example of the phenalenyl π-dimer and its radical cation.[34]

In this work, we propose a simple self-interaction-free method to include dynamic correlation effects during orbital optimization that is intermediate in cost and complexity between Kohn-Sham DFT and optimized orbital (Brueckner) coupled cluster methods. The idea is to produce scaled opposite-spin MP2-type (SOS-MP2)[35] optimized orbitals,

4

i.e., we want to find the set of optimal orbitals that makes the energy, which now includes a scaled second order perturbative correction to the correlation effects (see Eq. (1) below), stationary with respect to orbital rotations. This method may be described as "orbital-optimized opposite-spin scaled second order correlation" which we abbreviate as O2.

Opposite-spin MP2 (OS-MP2) methods[35,36] empirically enlarge the MP2 expression for alpha-beta correlation, while the same-spin contribution is completely neglected. These simple models were motivated by the work of Grimme who proposed to scale the two spin-components of the MP2 correlation energy with different factors (SCS-MP2).[37] OS-MP2 methods can be partly justified based on the fact that the opposite-spin correlation effects are larger in magnitude, as same-spin electrons are already correlated in the mean-field reference due to Fermi statistics. They statistically improve upon MP2 theory for relative energies and other molecular properties.[35,36,38-44] OS-MP2 methods also offer computational advantages as their energies and analytical gradients[45] can be evaluated with only fourth order effort (without exploiting locality), unlike conventional MP2 (or SCS-MP2), which scales with the fifth power in system size. Further savings are possible when locality and sparsity effects are taken into consideration (e.g. see Jung et al.[46]).

With the inclusion of correlation effects during the orbital optimization procedure in the O2 method, we hope to capture most of the advantages of the Brueckner models (such as OD) at much reduced cost. For example, O2 ought to reduce spin contamination in the UHF reference to a large extent. Relative to Kohn-Sham methods, the O2 approach has the twin advantages of being self-interaction free (as it still uses a single-reference whose

Coulomb and exchange matrix elements are treated exactly), and naturally including dispersion interactions. Algorithmically, the O2 method can be easily incorporated into an existing iterative framework that determines the HF orbitals and energies using a convergence procedure like geometric direct minimization (GDM)[47] based on orbital gradients.

We report the working equations and implementation of the O2 model in Section 2. As a result of the orbital relaxation, correlation energies could be slightly overestimated when compared to SOS-MP2 energies. Hence we suspect that the optimal value of the scaling factor may be slightly lower than the prescribed value for SOS-MP2$\left(c_{SOS} = 1.3\right)$.[35] We examine this with a statistical study of the atomization energies of the G2 database of 148 molecules[48,49] in Section 3. We also evaluate the performance of the O2 model for describing geometries and vibrational frequencies of a range of small doublet radicals and simple closed-shell molecules and compare its performance with other theories like HF, MP2 with the resolution-of-the-identity approximation (RI-MP2)[50,51] and SOS-MP2, and the more expensive OD method. In order to estimate the extent of spin contamination, we have also calculated the $\left\langle \hat{S}^2 \right\rangle_{O2}$ values for the doublet radicals. As a final example which is large enough to be difficult to treat via Brueckner methods like OD, we revisit the phenalenyl radical and its sigma dimer,[11] to estimate the improvement in spin contamination and energetics over unrestricted HF and MP2 theories.

II. Theory

A. The O2 model

For each iterative step, the O2 energy is simply given by:

$$E_{O2} = E_{ref} + c_{OS}E_{OS} \tag{1}$$

E_{ref} is the mean-field reference energy while the opposite spin correlation energy E_{OS} is given by [35,36]:

$$E_{OS} = -\sum_{\tau}^{N_\tau} w_\tau \sum_{PQ} X_{PQ}^\alpha(\tau) X_{PQ}^\beta(\tau) \tag{2}$$

In the SOS-MP2 and MOS-MP2 methods, Eq. (2) is the final working energy expression that is derived by utilizing a combination of auxiliary basis expansions and Laplace transformation to achieve fourth order computational scaling.[35,36] It can be written as:

$$X_{PQ}^\alpha = \sum_{ia}^\alpha B_{ia}^P(\tau) B_{ia}^Q(\tau) \tag{3}$$

where the **B** matrix is given in terms of two- and three-center integrals:

$$B_{ia}^P(\tau) = \sum_R (ia|R)(R|P)^{-\frac{1}{2}} \exp\left[(\varepsilon_i - \varepsilon_a)t_\tau\right] \tag{4}$$

In the above expressions, indices $i, j, \ldots$ are occupied orbitals while $a, b, \ldots$ refer to virtual orbitals, and the bar above the orbitals indicates β spin. $P, Q, \ldots$ indicate auxiliary basis functions while $\tau, w_\tau, t_\tau, N_\tau$ represent the quadrature point, the associated quadrature weight and root, the total number of quadrature points used, respectively. Further details about the implementation of Eqs. (3) and (4) is available in Refs. Jung et al.[35] and Lochan et al.[36] The optimal scaling factor c_{OS} used in Eq. (1) is determined in the next section through a statistical study.

To find the optimal set of orbitals, we require E_{O2} to satisfy the following stationary condition[26]:

$$\frac{dE_{O2}}{d\mathbf{U}}\frac{d\mathbf{U}}{d\theta} = 0 \tag{5}$$

where $\mathbf{U}$ is the unitary transformation matrix defined in terms of rotation angles θ that transforms the old set of orbitals into the new optimal ones, namely, $\mathbf{C}_{new} = \mathbf{C}_{old}\mathbf{U}$. The term $\frac{d\mathbf{U}}{d\theta}$ is independent of the model being used and is described elsewhere.[52] As the energy, E_{O2} is affected only by the virtual-occupied (vo) rotations, we only need to evaluate the orbital gradient: $\frac{dE_{O2}}{dU_{vo}}$, which can be easily expressed as:

$$\frac{dE_{O2}}{d\mathbf{U}} = 2F_{vo} + c_{OS} \cdot 2L_{vo} \tag{6}$$

In the above expression, the orbital derivative of the reference energy is simply given by the vo elements of the Fock matrix $\left(F_{vo}\right)$[47] while the OS Lagrangian $\left(L_{vo}\right)$ evaluated with the current basis gives the orbital gradient of the OS correlation energy.[26] The expression for L_{vo} retains the same form as the SOS-MP2/MOS-MP2 Lagrangian but includes an extra term $(L4)_{vo}$:

$$L_{vo} = \left(L1\right)_{vo} + \left(L2\right)_{vo} + \left(L3\right)_{vo} + \left(L4\right)_{vo} \tag{7}$$

The detailed expression for L1, L2, L3 are described elsewhere.[45,53] L4 arises because the orbitals do not satisfy the Brillouin condition $F_{vo} = 0$, and is given by [54]:

$$\left(L4\right)_{vo} = F_{vo}P_{oo} + F_{ov}P_{vv} \tag{8}$$

where, P_{oo}, P_{vv}, represent the occupied-occupied and virtual-virtual elements of the OS one-particle density matrix (P).[45] The computational cost of evaluating the Lagrangian, L_{ov} is roughly twice the cost of evaluating the OS correlation energy.

The correlation correction to both the energy in Eq. (1) and its orbital derivative in Eq. (6) can be readily incorporated into an orbital-gradient based SCF framework such as geomtric direct minimization (GDM).[47] The resulting procedure becomes:

1. Generate an initial guess set of orbitals: for instance generalized Wolfsberg-Helmholtz (GWH)[55], core Hamiltonian or HF optimized orbitals can be used.

2. Build the Fock Matrix.

3. Evaluate E_{OS}, L_{vo} with the current set of orbitals.

4. Update $E_{O2}, \dfrac{dE_{O2}}{d\mathbf{U}}$ for the current iteration.

5. Generate a new orbital step and thus a new set of molecular orbital (MO) coefficients through the GDM procedure.

6. If convergence (based on the maximum gradient component) is not achieved, go back to step 2.

We observe that the orbital gradient could alternatively be used to define a generalized Fock matrix as the basis for a more conventional diagonalization-based SCF procedure for the O2 model, but we shall not pursue this question further here.

B. Evaluation of $\left\langle \hat{S}^2 \right\rangle_{O2}$

The expectation value of total spin operator, $\hat{S}^2$, measures spin contamination in the wavefunction.[22,23,56,57] For methods like O2 without well-defined wavefunctions, we evaluate $\langle \hat{S}^2 \rangle$ as the response to a perturbation λS^2 that is added to the Hamiltonian:

$$\langle \hat{S}^2 \rangle = \frac{dE(\lambda)}{d\lambda} \Bigg|_{\lambda=0} \tag{9}$$

By applying standard perturbation theory we can express $\langle \hat{S}^2 \rangle_{O2}$ as [22]:

$$\langle \hat{S}^2 \rangle_{O2} = \langle \hat{S}^2 \rangle_{ref} + 2 \langle \phi_o | S^2 | \phi_1 \rangle \tag{10}$$

Here, ϕ_o, ϕ_1 represent the reference and the first order O2 correction to the wavefunction. As only doubly excited determinants contribute to ϕ_1, we can re-express Eq. (10) after a few steps of algebra as:

$$\langle \hat{S}^2 \rangle_{O2} = \langle \hat{S}^2 \rangle_{ref} + c_{OS} \sum_{ia}^{\alpha} \sum_{jb}^{\beta} \frac{S_{i\bar{b}} S_{\bar{j}a} (ia | \overline{jb})}{\Delta_{ia}^{\overline{jb}}} \tag{11}$$

Here, $S_{i\bar{b}} = \int \phi_{i_+} s_{\bar{b}} d\tau$ and $S_{\bar{j}a} = \int \phi_{\bar{j}} s_{-} \phi_a d\tau$ [23] and $(ia | \overline{jb})$ are the 4-center 2-electron integrals in Mulliken notation. The energy denominator is $\Delta_{ia}^{\overline{jb}} = \varepsilon_a + \varepsilon_{\bar{b}} - \varepsilon_i - \varepsilon_{\bar{j}}$, where ε_p is the orbital energy. $\langle \hat{S}^2 \rangle_{ref} = \left[s(s+1) + n_\beta \right] - \sum_i^{\alpha} \sum_j^{\beta} S_{i\bar{j}} S_{\bar{j}i}$, where n_β is the number of β spin electrons.[23]

We now introduce auxiliary basis functions and the Laplace scheme to avoid the formal 5th order cost of evaluating Eq. (11) (in making the $(ia | \overline{jb})$ integrals) in a procedure similar to that followed in SOS-MP2[35] to get,

$$\left\langle \hat{S}^2 \right\rangle_{O2} = \left\langle \hat{S}^2 \right\rangle_{ref} + c_{OS} \sum_{\tau}^{N_{\tau}} w_{\tau} \sum_{ia}^{\alpha} \sum_{jb}^{\beta} \sum_{P} S_{i\bar{b}} S_{\bar{j}a} B_{ia}^{P}(\tau) B_{\bar{j}\bar{b}}^{P}(\tau) \qquad (12)$$

The product of the two Laplace-transformed **B** matrices (Eq. (4)) approximately represents the fraction of the four-center two-electron integral over the energy denominator in Eq. (11). The evaluation of this expression now scales as $O(N_{\tau}.o^2vX)$, where, o, v, X represent the number of occupied, virtual and auxiliary basis functions, respectively. This is much smaller than the dominant step in O2 energy evaluation, namely Eq. (3), which requires about $O(N_{\tau}.ovX^2)$ operations (since $X \approx 3N_{bas}$ (number of basis functions), and $N_{bas} \approx v > o$). We note here that Eq. (12) also corresponds to the $\left\langle \hat{S}^2 \right\rangle$ expression for RI-MP2 and SOS-MP2, with $c_{OS} = 1$ and $c_{OS} = 1.3$, respectively.

III. Results and Discussion

We have implemented the O2 model described above in a developmental version of QCHEM.[58] We determine the optimal scaling factor c_{OS} used in Eq. (1) by analyzing the performance of the O2 method relative to experiment in a statistical study of various properties like atomization energies, bond lengths and vibrational frequencies of both closed- and open-shell systems. As the O2 analytical gradients and Hessians are not yet available, the optimized geometries and vibrational data at this level are obtained by applying the standard finite difference procedure to the O2 energy and gradient, respectively. The OD, HF, RI-MP2 and SOS-MP2 optimized geometries are obtained using their respective analytical gradients, while the frequencies are obtained by the finite difference of their gradients. In this report, we use unrestricted HF orbitals as the initial

reference for all the calculations and apply constraints to ensure that the alpha-beta spin symmetry is broken for closed-shell systems in order to locate any possible unrestricted solution. A tight integral threshold of 10^{-14} was employed and the SCF convergence criterion was set to 10^{-8}. Also, all electrons were correlated for the reported molecules.

A. Atomization energies

The 148 small molecules of the extended G2 database[48,49] consists of 118 closed-shell molecules and 30 open-shell systems. We have evaluated the atomization energy using the O2 model for various c_{OS} values (between 0.9-1.4) using different standard basis sets of either triple-zeta or quadruple-zeta quality and corresponding auxiliary basis sets.[59] Table 1 summarizes the statistical errors (MSE: mean signed error; MAE: mean absolute error; RMS: root mean squared error; MAX: maximum absolute error) obtained relative to "experimental electronic" atomization energies for the various scaling factors and basis sets used. The experimental values are obtained from Ref. 60 where the zero point energy is subtracted from the original experimental data to allow direct comparison with the estimated electronic energies from the indicated methods.

The triple-zeta quality basis, both Pople-type (6-311G(2df,2pd))[61,62] and the Dunning basis set (cc-pVTZ),[63] predict that the statistical errors relative to experiment are minimal for the scaling factor $c_{os} = 1.3$. The RMS error is about 5 kcal/mol while the MSE is about -1 to -1.6 kcal/mol indicating that the atomization energies are almost equally overestimated and underestimated across the 148 molecules. The MAX error of ~ 21 kcal/mol is encouraging, and in fact, a closer look at the absolute errors indicated that only 9 molecules in the whole set had errors above 10 kcal/mol. The atomization energies

seem to be quite sensitive to the scaling factor as increasing or decreasing the scaling factor by 0.1 units increases the RMS error by almost a factor of 2-3. This shows both the advantage as well as the potential disadvantage of having an empirical parameter in the model.

As mentioned in the Introduction, we expect the optimal scaling factor for the O2 method to be slightly lower than the prescribed SOS-MP2 value of $c_{SOS} = 1.3$[35] due to orbital relaxation, which leads to overestimation of the correlation energy. However, we do not observe this here, because in the comparison against experiment with a triple-zeta quality basis we are trying to compensate for both limitations of the correlation treatment and basis set incompleteness effects. The latter necessarily depends on the basis used. To study the effect of the basis set quality on the atomization energies and the scaling factors, we have obtained the statistical errors for a quadruple-zeta quality Dunning basis set (cc-pVQZ).[64] Since the Dunning basis sets are systematically constructed, we can estimate the complete basis set (CBS) limit by the two-point extrapolation scheme suggested by Halkier et al. [65]

The cc-pVQZ and the T-Q extrapolated CBS limit (termed as cc-pV(TQ)Z) results are also indicated in Table 1. At the cc-pVQZ level, the $c_{OS} = 0.9 - 1.2$ results are improved relative to the cc-pVTZ results. In particular, the $c_{OS} = 1.3$ results are relatively worse with cc-pVQZ than with cc-pVTZ: the MAE and RMS errors have almost doubled. At the same time, the quality of the $c_{OS} = 1.2$ results is improved considerably. The MSE, MAE and RMS errors are almost halved compared to the corresponding cc-pVTZ case, while the MAX error decreases by almost 3.5 kcal/mol relative to $c_{OS} = 1.3$/cc-pVTZ. This improvement in results with the shift in the scaling factor is consistent with our

earlier belief that the optimal scaling factor must be slightly lower than the prescribed value for SOS-MP2 due to the orbital relaxation effects. This notion is further confirmed by looking at the T-Q extrapolated CBS limit results. We now have a new soft minimum at $c_{OS} = 1.2$ with respect to all the statistical errors considered. The RMS error is about 5.6 kcal/mol, while the MAX error is about 18 kcal/mol. Therefore our best estimates of the atomization energies indicate that the optimal scaling factor for predicting relative energies is about $c_{OS} = 1.2$, as we approach the basis set limit. For smaller basis sets, larger factors are optimal as already discussed previously.

In SOS-MP2 and MOS-MP2 theories,[35,36] the quality of the method with respect to the scaling factor was determined by comparison against a more advanced theory like QCISD(T)[66] for a given basis set. This also minimizes basis-set specific effects by assessing the quality of the given method against the best available estimate of that particular property within a given basis set.[67] In the same spirit, we have also compared the O2 atomization energies against the corresponding QCISD(T) results calculated with the cc-pVTZ basis. The resulting errors are shown in Table 2 for various scaling factors. We find that the $c_{OS} = 1.2$ case indeed gives the least errors relative to QCISD(T), consistent with the estimated CBS limit results in Table 1. We have also included the statistical errors relative to QCISD(T) from our previous work on SOS-MP2[35] and MOS-MP2[36] to compare against our O2 ($c_{OS} = 1.2$) results in Table 2. These results indicate that the MP2-type methods provide tremendous improvement over HF theory in general, and scaled MP2 methods improve upon MP2 itself. The quality of the O2 method with $c_{OS} = 1.2$ is very comparable (and slightly better than in terms of RMS and MAE errors) to SOS-MP2 and MOS-MP2.

B. Bond lengths

We will now examine the performance of the O2 model in describing the structural features of 12 simple first-row doublet radicals (BO, CF, CH, CN, CO^+, FH^+, NO, OF, OH, F_2^+, N_2^+, O_2^+) considered by Beran et al.[31] and 17 small, simple closed shell systems (corresponding to 24 bond lengths) CH_2O, $^1CH_2$, CH_4, CO, CO_2, F_2, H_2, H_2O, HCCH, HCN, HF, HNC, HOF, HOOH, N_2, N_2H_2, NH_3) using the 6-311G(2df,2pd) basis set. The experimental bond lengths of these molecules were obtained from Ref. 31, 68, 4 and references therein. Figures 1 and 2 show the plot of the errors ($= r_{Method} - r_{Expt}$) obtained in the bond-lengths relative to experiment by unrestricted HF, OD, RI-MP2, SOS-MP2 and O2 ($c_{OS} = 0.9 - 1.3$) methods for the radicals and closed-shell systems, respectively. The details of the obtained errors by the indicated methods are shown explicitly for each molecule in Table S1 and S2 (in supporting information). Table 3 summarizes the net statistical errors obtained for both types of system considered here.

Figures 1 and 2 clearly reflect the well-known tendency of HF theory to underestimate bond lengths for both open- and closed- shell systems due the absence of dynamic correlation effects (which slightly depopulate bonding orbitals, and slightly populate antibonding orbitals). The RMS error for HF is about 4.3 pm and 3.3 pm, respectively, for each case. OD also shows a tendency to underestimate bond distances (presumably due to neglect of connected triple excitations and beyond) although to a much lesser extent than HF. OD shows RMS errors of only about 1.5 pm and 0.9 pm, respectively for the radicals and closed-shell molecules. RIMP2 and SOS-MP2 slightly improve the HF description of the radical bond lengths by bringing down the HF RMS

error to about 3 pm and 2.7 pm, respectively, but clearly inclusion of second order correlation beyond HF does not help in reliably predicting accurate radical bond distances. The RMS, MAE and MAX errors of SOS-MP2 can be decreased by almost 50% by employing optimized O2 orbitals. However, the RI-MP2 and SOS-MP2 optimized bond lengths for the closed-shell systems are almost as accurate as the OD description. In fact, SOS-MP2 appears to predict geometries that are slightly better than OD! Also, the statistical results obtained for SOS-MP2 are in close correspondence with that of O2 ($c_{OS} = 1 - 1.2$), which again confirms the notion that the optimal OS scaling factors for O2 should be slightly lower than the prescribed $c_{OS} = 1.3$ for SOS-MP2.

For the smallest scaling factor considered here ($c_{OS} = 0.9$), O2 displays a tendency to underestimate the radical bond lengths falling very close to the OD curve, while $c_{OS} = 1.2, 1.3$ tend to predict longer bonds. The MSE errors shown in Table 3 further indicate that the largest cancellation of errors occurs with $c_{OS} = 1.0, 1.1$ for the radicals where the RMS errors are about 0.6 pm in each case. A similar trend is seen for the closed-shell molecules, except that the optimal scaling factors are shifted slightly towards $c_{OS} = 1.1, 1.2$. Also, the closed-shell systems appear to be rather insensitive towards the scaling factor as the RMS error varies only between 0.5-1.0 pm.

For both OD and HF, the most significant outlier amongst the open-shell systems is F_2^+ (absolute error of 3.9 pm and 11.1 pm, respectively), followed by O_2^+, where the errors are 1.1 pm and 5.5 pm respectively. O2 ($c_{OS} = 1.1$) manages to reduce this error to about 0.5 pm and 0.9 pm, respectively, for both molecules. For RI-MP2 and SOS-MP2, the largest outliers correspond to N_2^+ (absolute error of 7 pm and 6 pm, respectively) and

CN (absolute error of about 5 pm each), while O2 ($c_{OS} = 1.1$) decreases this error to about

1 pm. For closed-shell systems, the F-F, F-O and O-O bonds in F_2, HOF and HOOH

molecules present the three largest deviations from experiment for both HF and OD with

MAX errors of 8.4 pm and 2.2 pm, respectively, for F_2. The O2 model with $c_{OS} = 1.1, 1.2$

reduces this absolute error in the F-F bond length to about 0.4 pm and 0.9 pm,

respectively. The MAX error for these two scaling factors occurs with the singlet CH_2

system with 1.8 pm and 1.6 pm, respectively, which is in close agreement with the

corresponding MAX errors obtained with RI-MP2 and SOS-MP2. The combined

statistics of the 29 molecules considered in this study are also shown in Table 3, which

reflects that $c_{OS} = 1, 1.1$ produce the least errors. However, the performance of $c_{OS} = 1.2$

is only slightly poorer, with MAE, RMS and MAX errors for the combined set that are

comparable to the OD errors and provide significant improvement over not only the HF

description of both open- and closed-shell systems, but also the MP2 description of

radical bond lengths.

C. Vibrational Harmonic frequencies

The harmonic vibrational frequencies of the same 29 molecules used above were

determined at the respective HF, OD, RI-MP2, SOS-MP2 and O2 ($c_{OS} = 0.9 - 1.3$)

optimized geometries using the 6-311G(2df,2pd) basis set. The dataset thus consists of 12

frequencies from the simple doublet radicals and 54 frequencies from the 17 closed-shell

molecules. Figures 3 and 4 display the estimated errors ($= \omega_{Method} - \omega_{Expt}$) relative to

experimental data for the open- and closed-shell systems, respectively. In Figure 4, for

systems containing more than 2 atoms, only the two largest absolute errors are shown in

the plot for clarity. The details of the obtained errors by the indicated methods are shown explicitly for each molecule in Table S3 and S4 (in supporting information). The experimental harmonic frequencies were acquired from Ref. 6,31,69,70 and references therein for all molecules except $^1CH_2$, HOF, HOOH, N_2H_2. For the latter four molecules, the experimental frequencies were obtained from Ref. 71. Table 4 lists the statistical errors obtained across the radicals set and the closed-shell systems.

Figures 3 and 4 corroborate the conventional wisdom that the frequencies predicted by HF are too high with respect to the experiment. This is quite natural, as the HF method tends to shorten the bonds. Consistent with previous studies,[6] the MP2-type description of the radical bond-frequencies is very erratic, with RMS errors of 540 cm^{-1} and 525 cm^{-1}, respectively, for RI-MP2 and SOS-MP2, which is higher than the corresponding RMS error obtained for HF theory. With the inclusion of correlation during orbital optimization, both the OD and the O2 models are able to reduce the extent of these errors in the vibrational frequencies. The high level OD treatment achieves RMS and MAE errors of about 99 cm^{-1} and 92 cm^{-1}, respectively, for the radicals. With O2 ($c_{OS} = 1.2, 1.3$), the dynamic correlation is overestimated leading to longer bonds and smaller frequencies. The deviation from experiment is quite controlled in case of O2 ($c_{OS} = 0.9 - 1.1$), with RMS errors ranging between 65-87 cm^{-1} for the radicals and is consistently better than OD. This again suggests that for a particular property, the scaling parameter can be fine tuned to achieve good accuracy with respect to experimental data.

As shown in Table 4, the statistical errors for the radicals indicate that the best performance is obtained by O2 ($c_{OS} = 1.0$), with an RMS error of about 65 cm^{-1}, which is almost 30 cm^{-1} lower than the corresponding OD result. The largest deviation of the HF

and O2 ($c_{OS} = 1.1 - 1.3$) frequencies from experiment is observed with the molecules, F_2^+ and O_2^+, as for the bond-lengths. In this case, the absolute error for O_2^+ is decreased from about 590 cm^{-1} obtained with HF to about 40 cm^{-1} with O2 ($c_{OS} = 1.0$). The corresponding absolute error obtained with O2 ($c_{OS} = 1.2$) is about 260 cm^{-1}. RI-MP2 and SOS-MP2 produce disastrous results for NO, CN and CO^+, with absolute errors on the order of 1300 cm^{-1}, 900 cm^{-1} and 800 cm^{-1}, respectively, while the corresponding O2 ($c_{OS} = 1.2$) errors are less than 100 cm^{-1}.

With the closed-shell systems, Figure 4 indicates that the lower scaling factors, $c_{OS} = 0.9, 1$, display a tendency to predict higher vibrational frequencies, although the extent of the overestimation is much lower than that of HF. Also, the RI-MP2 and SOS-MP2 description of the vibrational frequencies is very good with MAE errors on the order of 60-70 cm^{-1} and their description appears to be better than OD (MAE is about 87 cm^{-1}). In particular, the obtained statistical errors for SOS-MP2 are in close correspondence with those of O2 ($c_{OS} = 1.2$). The absolute errors corresponding to O2 ($c_{OS} = 1.1, 1.2$) lie within 100 cm^{-1} except for a few prominent outliers. From Figure 4, the major outliers with errors greater than 400 cm^{-1} correspond to the frequencies of $^1CH_2$, HOF, HOOH, N_2H_2 for HF theory and correspondingly, high errors greater than or close to 200 cm^{-1} are uniformly associated with the OD, RI-MP2, SOS-MP2 and O2 model for all the scaling factors considered. As mentioned earlier, the experimental data corresponding to these molecules do not represent pure harmonic frequencies. Consequently, the lack of anharmonic corrections to the theoretical results produces errors that are too high.

As these errors due partly to anharmonicity would bias our analysis, we have excluded these four molecules from our statistical study of the closed-shell systems and shown the corresponding results in Table 4. Following this, the RMS, MAE and MSE values decrease by almost 30-50% for all the O2 cases, the MP2-type methods and OD, and by about 20% for HF. The MAX error for all the O2 cases and HF drops by at least 100 cm^{-1} and 170 cm^{-1}, respectively, while the RI-MP2 and SOS-MP2 MAX errors decrease by almost 50-60% and the OD MAX error is about four times smaller. The statistical errors for the modified closed-shell set clearly indicates that the RI-MP2 and SOS-MP2 vibrational frequency results are on par with OD, with SOS-MP2 providing a marginally improved performance that is quite close to the O2 ($c_{OS} = 1.2$) results. The new significant outliers are N_2 and F_2 for HF theory with absolute errors on the order of 350-375 cm^{-1}. With the O2 model ($c_{OS} = 1.1, 1.2$), these errors can be decreased by almost 93% for N_2 and by about 75-90% for F_2. The statistical errors in Table 4 corresponding to the modified closed-shell set now indicate that $c_{OS} = 1.2$ is the optimal scaling factor.

Table 4 also reports the combined statistics obtained for all the open-shell and closed-shell molecules (excluding $^1CH_2$, HOF, HOOH, N_2H_2). This modified set now includes 48 harmonic vibrational frequency data. Table 4 indicates that the O2 model with any of the scaling factors considered here can obtain significant improvement over HF, RI-MP2 and SOS-MP2 theory as far as predicting vibrational frequencies goes and O2 ($c_{OS} = 1.0, 1.1$) provide OD quality results. The RMS and MAE values for O2 ($c_{OS} = 0.9 - 1.2$) are within 73 ± 8 cm^{-1} and 59 ± 9 cm^{-1}, respectively, indicating that changes of about 0.1 units in the scaling factors are quite insensitive towards the overall performance of the model. Therefore, in keeping with our previous analysis on

atomization energies, bond lengths and now vibrational frequencies, we conclude that the scaling factor for the O2 model that gives most reasonable performance across all properties we've considered is likely to be $c_{OS} = 1.2$. We retain this optimal scaling factor for the remaining discussion in this paper.

D. Spin contamination

The evaluation of $\langle \hat{S}^2 \rangle$ is often used as a diagnostic tool to estimate the extent of spin contamination present in the wavefunction.[56,57] Using the procedure described in the Theory section, we have calculated and reported the $\langle \hat{S}^2 \rangle$ values in Table 5 for the O2 model with $c_{OS} = 1.2$ for the 12 doublet radicals considered in the preceding section and compared it to the corresponding estimates from HF, OD, RI-MP2 and SOS-MP2 theories with the 6-311G(2df,2pd) basis set. The extent of spin contamination at the HF level of theory is quite high for radicals like CN and CO^+, where the predicted $\langle \hat{S}^2 \rangle$ value is 1.156 and 1.003, respectively, as against the actual value of 0.75 for a doublet spin state. With the inclusion of correlation effects in RI-MP2, this decreases slightly to 1.150 and 0.968, respectively. With the O2 method, this improves to 0.752 and 0.755, respectively, which is almost comparable to the performance of the more advanced (and far more computationally demanding) OD treatment. Overall, the mean absolute percentage error (MAPE) is decreased from almost 7% (HF theory) to a mere 0.2% at the O2 level, and the maximum deviation from the actual value of 0.75 is only 0.005. The MAPE of RIMP2 and SOS-MP2 (about 6% each) indicate that only marginal

improvements can be obtained with the inclusion of second order correlation beyond the HF description.

The effect of spin contamination in the HF wavefunction sometimes worsens as the radical gets bigger in size. We will now re-examine such an example where the effect of high-spin contamination can be correlated with poor relative energies. The phenalenyl doublet π-radical ($P^\bullet$) can associate to form π-stacked[34] or σ-type (P_2^σ)[11] dimers. It was the subject of a recent computational study where it was found that the unrestricted HF $\left\langle \hat{S}^2 \right\rangle$ values for the $P^\bullet$ radical and the various P_2^σ dimers were as high as 2.1 and 3.2, respectively, in place of the actual values of 0.75 and 0.[11] Also it was found that unrestricted HF theory did not bind (BE = 31 kcal/mol) the σ-dimer while the corresponding unrestricted MP2 method predicted very low interaction energies (BE = -6.0 kcal/mol for the RR1 isomer) when the binding energy (prior to basis set super position error (BSSE) or zero-point energy corrections) was simply calculated as:

$$BE = E_{P_2^\sigma} - 2E_{P^\bullet} \qquad (13)$$

The authors had to resort to an alternative scheme that split the interaction in terms of an isogyric and non-isogyric process in order to get a good consistent description. As the latter step involved smaller molecules, it was treated with a more sophisticated computational method such as CCSD(T) while the former step was treated with HF and MP2, which are known to perform well with isodesimic/isogyric processes.[11]

This is the type of case where our new O2 method may exhibit its advantages over HF and MP2 theory, which perform poorly, and OD theory, which is computationally too expensive to be applied to this system. We have calculated the $\left\langle \hat{S}^2 \right\rangle$ value and the

binding energy calculated according to Eq. (13) for the RR1 isomer of the P_2^{σ}-dimer at the unrestricted HF, RI-MP2 and O2 ($c_{OS} = 1.2$) levels of theory using 6-31G* basis. The UHF/6-31G* optimized structures of the respective species were taken from Small et al.[11] The results shown in Table 6 indicate that the O2 model is able to almost eliminate spin contamination in the $P^{\bullet}$ radical. In this case, the $\langle \hat{S}^2 \rangle$ value decreases from 2.1 at the HF level to about 0.76. For the P_2^{σ}-dimer, O2 is able to locate the correct restricted solution, thereby completely eliminating the spin-contamination found at the HF level. We emphasize here that the respective O2 calculations on the dimer (spin state = 0) were performed using the unrestricted HF reference orbitals and constraints were applied to ensure that the alpha-beta spin symmetry is broken in order to find any possible unrestricted solution. Also note that RI-MP2 only manages to improve $\langle \hat{S}^2 \rangle_{HF}$ by about 0.2-0.3 units for both $P^{\bullet}$ and its sigma dimer.

The resulting binding energy evaluated according to Eq. (13) at the O2 level of theory is about -21.6 kcal/mol for the RR1 dimer, which is consistent with the corresponding result of -20.8 kcal/mol evaluated using the alternative scheme proposed by Small et al.[11] Therefore, the O2 model offers a simple and relatively cheap way to clean up the spin-contamination problems and predict correct energetics in a straightforward manner for complex systems like phenalenyl dimers that are computationally too expensive to be treated at the OD level, for instance.

IV.   **Conclusions**

We have extended the concept of using scaled opposite-spin second order correlation energy to obtain optimal orbitals that effectively improve the HF description of both open- and closed-shell systems in terms of energetics and geometrical features. We call this simple, second order variant of the optimized orbital coupled cluster doubles (OD) method, "orbital-optimized opposite-spin scaled second order correlation" theory (O2). Without cutoffs or use of sparsity, the O2 energy can be evaluated with computational effort that scales only as the 4th power of system size (with cubic disk and quadratic memory), unlike OD theory, which requires 6th order computation, 4th order disk storage.

We recommend using 1.2 as the optimal opposite-spin scaling factor in the O2 model based on the results of a statistical study (using experimental data as reference) of atomization energies, bond lengths and vibrational frequencies of small closed-shell systems and doublet radicals. This study also indicated that OD-type results could be obtained for describing the bond-distances and vibrational frequencies of small doublet radicals and the HF harmonic vibrational frequencies could be improved by almost 70-80%. While the O2 description of closed-shell systems is probably as good as and sometimes even better than SOS-MP2, in case of open-shell systems, the O2 model outperforms both SOS-MP2 and RI-MP2.

We also found that small changes to the magnitude of the optimal scaling factor do not greatly affect the performance of the O2 model in predicting molecular geometry and frequencies. However, the magnitude of the chosen scaling factor quite sensitively affects atomization energies. The need to empirically choose the parameter value is therefore one of the biggest disadvantages of the O2 model, which is true of all opposite-spin scaled

techniques.[35,36] However, for particular chemical applications, this may be turned to advantage, by tuning its value for the problem at hand.

In this paper, we also show that one of the significant advantages of the O2 model is its ability to almost eliminate the spin-contamination problem of the unrestricted HF reference wavefunction. We further demonstrated that the O2 model could clean up the large spin-contamination of the phenalenyl radical and its sigma-dimer at the UHF level and obtain binding energies that are consistent with higher level methods. We therefore believe the O2 method is a promising technique for treating large radicals- it bridges the gap between optimized orbital coupled cluster doubles methods (OD, BD etc) that are too computationally expensive, while not suffering from deficiencies of present-day density functional theory methods such as self-interaction errors and neglect of dispersion forces.

V. Acknowledgement

VI. Supporting Information

The experimental bond-lengths and vibrational frequencies used for the statistical study of the 29 molecules and the respective errors obtained relative to experimental data by the HF, OD, RIMP2, SOSMP2 and O2 methods are given.

VII. References

[1]   C. Moller and M. S. Plesset, Phys. Rev. **46** (7), 0618 (1934).

[2]   G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76** (4), 1910 (1982).

[3]   K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157** (6), 479 (1989).

[4]   T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular electronic structure theory*. (John Wiley & Sons Ltd., New York, 2000).

[5]   F. Jensen, *Introduction to computational chemistry*. (John Wiley & Sons, Inc., New York, 1999).

[6]   E. F. C. Byrd, C. D. Sherrill, and M. Head-Gordon, J. Phys. Chem. A **105** (42), 9736 (2001).

[7]   L. Farnell, J. A. Pople, and L. Radom, J. Phys. Chem. **87** (1), 79 (1983).

[8]   D. J. Tozer, N. C. Handy, R. D. Amos, J. A. Pople, R. H. Nobes, Y. M. Xie, and H. F. Schaefer, Mol. Phys. **79** (4), 777 (1993).

[9]   M. W. Wong and L. Radom, J. Phys. Chem. A **102** (12), 2237 (1998).

[10]   F. Jensen, Chem. Phys. Lett. **169** (6), 519 (1990).

[11]   D. Small, S. V. Rosokha, J. K. Kochi, and M. Head-Gordon, J. Phys. Chem. A **109** (49), 11261 (2005).

[12]   P. M. W. Gill, J. A. Pople, L. Radom, and R. H. Nobes, J. Chem. Phys. **89** (12), 7307 (1988).

[13]   P. M. W. Gill and L. Radom, Chem. Phys. Lett. **132** (1), 16 (1986).

[14]   R. H. Nobes, J. A. Pople, L. Radom, N. C. Handy, and P. J. Knowles, Chem. Phys. Lett. **138** (5), 481 (1987).

[15] A. I. Krylov, C. D. Sherrill, E. F. C. Byrd, and M. Head-Gordon, J. Chem. Phys. **109** (24), 10669 (1998).

[16] R. B. Murphy and R. P. Messmer, Chem. Phys. Lett. **183** (5), 443 (1991).

[17] R. D. Amos, J. S. Andrews, N. C. Handy, and P. J. Knowles, Chem. Phys. Lett. **185** (3-4), 256 (1991).

[18] P. J. Knowles, J. S. Andrews, R. D. Amos, N. C. Handy, and J. A. Pople, Chem. Phys. Lett. **186** (2-3), 130 (1991).

[19] M. Rittby and R. J. Bartlett, J. Phys. Chem. **92** (11), 3033 (1988).

[20] P. O. Lowdin, Phys. Rev. **97** (6), 1509 (1955).

[21] A. T. Amos and G. G. Hall, Proc. R. Soc. London, A **263** (131), 483 (1961).

[22] H. B. Schlegel, J. Chem. Phys. **84** (8), 4530 (1986).

[23] W. Chen and H. B. Schlegel, J. Chem. Phys. **101** (7), 5957 (1994).

[24] R. A. Chiles and C. E. Dykstra, J. Chem. Phys. **74** (8), 4544 (1981).

[25] N. C. Handy, J. A. Pople, M. Headgordon, K. Raghavachari, and G. W. Trucks, Chem. Phys. Lett. **164** (2-3), 185 (1989).

[26] C. D. Sherrill, A. I. Krylov, E. F. C. Byrd, and M. Head-Gordon, J. Chem. Phys. **109** (11), 4171 (1998).

[27] W. Kohn and L. J. Sham, Phys. Rev. **140** (4A), 1133 (1965).

[28] R. G. Parr and W. Yang, *Density-functional theory of atoms and molecules*. (Oxford, New York, 1989).

[29] J. Baker, A. Scheiner, and J. Andzelm, Chem. Phys. Lett. **216** (3-6), 380 (1993).

[30] C. D. Sherrill, M. S. Lee, and M. Head-Gordon, Chem. Phys. Lett. **302** (5-6), 425 (1999).

[31] G. J. O. Beran, S. R. Gwaltney, and M. Head-Gordon, Phys. Chem. Chem. Phys. **5** (12), 2488 (2003).

[32] B. G. Johnson, C. A. Gonzales, P. M. W. Gill, and J. A. Pople, Chem. Phys. Lett. **221** (1-2), 100 (1994).

[33] S. Kristyan and P. Pulay, Chem. Phys. Lett. **229** (3), 175 (1994).

[34] D. Small, V. Zaitsev, Y. S. Jung, S. V. Rosokha, M. Head-Gordon, and J. K. Kochi, J. Am. Chem. Soc. **126** (42), 13850 (2004).

[35] Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, J. Chem. Phys. **121** (20), 9793 (2004).

[36] R. C. Lochan, Y. Jung, and M. Head-Gordon, J. Phys. Chem. A **109** (33), 7598 (2005).

[37] S. Grimme, J. Chem. Phys. **118** (20), 9095 (2003).

[38] M. Gerenkamp and S. Grimme, Chem. Phys. Lett. **392** (1-3), 229 (2004).

[39] M. Piacenza and S. Grimme, J. Comput. Chem. **25** (1), 83 (2004).

[40] T. P. M. Goumans, A. W. Ehlers, K. Lammertsma, E. U. Wurthwein, and S. Grimme, Chem.--Eur. J. **10** (24), 6468 (2004).

[41] S. Grimme, J. Phys. Chem. A **109** (13), 3067 (2005).

[42] S. Grimme, Chem.--Eur. J. **10** (14), 3423 (2004).

[43] S. Grimme, C. Muck-Lichtenfeld, E. U. Wurthwein, A. W. Ehlers, T. P. M. Goumans, and K. Lammertsma, J. Phys. Chem. A **110** (8), 2583 (2006).

[44] I. Hyla-Kryspin and S. Grimme, Organometallics **23** (23), 5581 (2004).

[45] R. C. Lochan, Y. Shao, and M. Head-Gordon, submitted (2006).

[46] Y. Jung and M. Head-Gordon, Phys. Chem. Chem. Phys. **8** (24), 2831 (2006).

[47]   T. Van Voorhis and M. Head-Gordon, Mol. Phys. **100** (11), 1713 (2002).

[48]   L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **106** (3), 1063 (1997).

[49]   L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94** (11), 7221 (1991).

[50]   M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208** (5-6), 359 (1993).

[51]   F. Weigend, M. Haser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294** (1-3), 143 (1998).

[52]   M. Head-Gordon and J. A. Pople, J. Phys. Chem. **92** (11), 3063 (1988).

[53]   R. A. DiStasio, R. P. Steele, Y. M. Rhee, Y. Shao, and M. Head-Gordon, *J. Comput. Chem.* in press (2006).

[54]   T. Van Voorhis and M. Head-Gordon, J. Chem. Phys. **117** (20), 9190 (2002).

[55]   M. Wolfsberg and L. Helmholz, J. Chem. Phys. **20** (5), 837 (1952).

[56]   A. I. Krylov, J. Chem. Phys. **113** (15), 6052 (2000).

[57]   H. Yuan and D. Cremer, Chem. Phys. Lett. **324** (5-6), 389 (2000).

[58]   Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. Distasio Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A.

M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, E. R. J. Ritchie, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Phys. Chem. Chem. Phys. **8**, 3172 (2006).

[59]   F. Weigend, A. Kohn, and C. Hattig, J. Chem. Phys. **116** (8), 3175 (2002).

[60]   M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110** (11), 5029 (1999).

[61]   R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72** (1), 650 (1980).

[62]   A. D. Mclean and G. S. Chandler, J. Chem. Phys. **72** (10), 5639 (1980).

[63]   R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96** (9), 6796 (1992).

[64]   T. H. Dunning, J. Chem. Phys. **90** (2), 1007 (1989).

[65]   A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286** (3-4), 243 (1998).

[66]   J. A. Pople, M. Head-Gordon, and K. Raghavachari, J. Chem. Phys. **87** (10), 5968 (1987).

[67]   M. Head-Gordon, J. Phys. Chem. **100** (31), 13213 (1996).

[68]   *CRC Handbook of Chemistry and Physics*, 76th ed. (Chemical Rubber Publishing Company, Boca Raton, 1995).

[69]   J. M. L. Martin, Chem. Phys. Lett. **242** (3), 343 (1995).

[70]   W. J. Hehre, L. Radom, P. V. Schleyer, and J. A. Pople, *Ab initio molecular orbital theory*, 1 ed. (John Wiley & Sons, Inc., New York, 1986).

[71]   *NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101*. (Aug 2005).

List of Figures

Figure 1. Plot of the deviation of the OD, HF, RIMP2, SOSMP2 and O2 ($c_{OS} = 0.9 - 1.3$) optimized bond lengths from the respective experimental values for the 12 doublet radicals using 6-311G(2df,2pd) basis set.

Figure 2. Plot of the deviation of the OD, HF, RIMP2, SOSMP2 and O2 ($c_{OS} = 0.9 - 1.3$) optimized bond lengths from the respective experimental values for the 17 closed-shell molecules using 6-311G(2df,2pd) basis set.

Figure 3. Plot of the deviation of the OD, HF, RIMP2, SOSMP2 and O2 ($c_{OS} = 0.9 - 1.3$) harmonic vibrational frequencies from the respective experimental values for the 12 doublet radicals using 6-311G(2df,2pd) basis set.

Figure 4. Plot of the deviation of the OD, HF, RI-MP2, SOS-MP2 and O2 ($c_{OS} = 0.9 - 1.3$) vibrational frequencies (only the top two largest errors are shown for all polyatomic molecules) from the respective experimental values for the 17 closed-shell molecules using 6-311G(2df,2pd) basis set.

Table 1. Calculated statistical errors[a] in atomization energies of the 148 G2 database molecules relative to experimental data using O2 method with c_{OS} varied between 0.9-1.4 and different basis sets (in kcal/mol).

Error	0.9	1.0	1.1	1.2	1.3	1.4
			6-311G(2df,2pd)			
MSE	-47.95	-36.42	-24.85	-13.23	-1.57	10.15
MAE	48.01	36.73	24.92	13.31	3.50	10.50
RMS	54.45	41.34	28.25	15.30	4.66	12.81
MAX	108.60	82.50	56.86	31.17	20.80	42.05
			cc-pVTZ			
MSE	-47.00	-34.84	-22.95	-11.03	0.95	12.97
MAE	47.06	35.14	23.02	11.12	3.72	13.37
RMS	52.79	39.15	25.73	12.59	5.17	16.44
MAX	101.10	73.82	46.49	25.81	21.28	42.66
			cc-pVQZ			
MSE	-42.38	-30.18	-17.93	-5.65	6.69	22.53
MAE	42.45	30.25	18.01	6.37	7.10	22.38
RMS	48.38	34.55	20.78	7.62	8.73	25.68
MAX	95.21	68.09	41.00	17.83	27.99	56.90
			cc-pV(TQ)Z[b]			
MSE	-40.17	-27.72	-15.23	-2.69	9.89	10.15
MAE	40.24	27.81	15.52	4.59	9.97	10.50
RMS	46.20	32.16	18.24	5.76	11.85	12.81
MAX	93.08	65.80	38.48	18.25	33.58	42.05

[a] MSE: mean signed error; MAE: mean absolute error; RMS: root mean squared error; MAX: maximum absolute error.

[b] Complete basis set limit estimated from the 2-point T-Q extrapolation scheme.

Table 2. Calculated statistical errors in the atomization energies relative to QCISD(T) values obtained by the various flavors of MP2 and the O2 method ($c_{OS} = 0.9 - 1.4$) with cc-pVTZ basis for the 148 molecules of the G2 database (in kcal/mol).

Error	0.9	1.0	1.1	1.2	1.3	1.4
MSE	-32.24	-20.17	-8.29	3.64	15.61	27.63
MAE	32.36	20.53	8.67	4.29	15.64	27.64
RMS	36.61	23.07	9.92	5.83	18.51	32.06
MAX	70.04	43.30	19.45	24.92	40.54	68.04

Error	HF[a]	RIMP2[a]	SCSMP2[a]	SOSMP2[a]	MOSMP2[a]	O2(1.2)
MSE	137.07	-10.19	-6.33	-4.34	2.80	3.64
MAE	137.09	11.68	6.84	5.00	4.91	4.29
RMS	156.51	14.87	8.42	6.18	6.18	5.83
MAX	312.15	38.12	24.43	21.08	16.51	24.92

[a] From Ref. 35,36.

Table 3. Statistical errors in bond lengths of 12 doublet radicals, 17 closed-shell systems and their combined set (29 molecules) relative to experimental bond lengths estimated at the OD, HF, RIMP2, SOSMP2 and O2 (c_{OS}=0.9-1.3) levels of theory using 6-311G(2df,2pd) basis (in Å).

Error	OD	HF	RIMP2	SOSMP2	0.9	1	1.1	1.2	1.3
				Doublet radicals					
MSE	-0.011	-0.032	-0.001	-0.005	-0.009	-0.003	0.002	0.008	0.018
MAE	0.011	0.034	0.023	0.019	0.009	0.005	0.006	0.011	0.021
RMS	0.015	0.043	0.03	0.027	0.012	0.006	0.006	0.014	0.032
MAX	0.039	0.111	0.07	0.061	0.033	0.016	0.012	0.035	0.098
				Closed-shell systems					
MSE	-0.006	-0.026	-0.002	-0.001	-0.008	-0.005	-0.002	0.001	0.005
MAE	0.006	0.026	0.002	0.003	0.008	0.005	0.003	0.004	0.007
RMS	0.009	0.033	0.009	0.005	0.01	0.007	0.005	0.006	0.01
MAX	0.022	0.084	0.021	0.018	0.025	0.019	0.018	0.016	0.024
				Combined Set					
MSE	-0.007	-0.028	-0.002	-0.002	-0.008	-0.004	-0.001	0.004	0.009
MAE	0.007	0.029	0.012	0.009	0.008	0.005	0.004	0.007	0.011
RMS	0.011	0.037	0.019	0.016	0.011	0.007	0.005	0.01	0.02
MAX	0.039	0.111	0.07	0.061	0.033	0.019	0.018	0.035	0.098

Table 4. Statistical errors in the harmonic vibrational frequencies of 12 doublet radicals, 17 closed-shell systems and their modified combined set[a] estimated at the OD, HF, RIMP2, SOSMP2 and O2 (c_{OS} = 0.9-1.3) levels of theory relative to experimental harmonic vibrational frequencies using 6-311G(2df,2pd) basis (in cm^{-1}).

Error	OD	HF	RIMP2	SOSMP2	0.9	1	1.1	1.2	1.3
Open-shell radicals									
MSE	91.8	243.2	205.5	245.8	66.3	27.9	-14.4	-64	-134
MAE	91.8	249.8	379.7	360.6	66.4	61.7	68.7	106.8	167.2
RMS	98.5	291.6	539.4	525.5	78.7	65.3	87.4	141.5	247.7
MAX	152.8	588.7	1345.7	1239.5	154.7	134.5	169.1	321.6	660.5
Closed-shell systems									
MSE	80.3	223	64.5	60.8	102.7	86.6	69.6	51.6	32.6
MAE	87.1	224.4	77.6	66.2	104.1	88.5	74.3	61.8	61.9
RMS	123.4	258.6	108.5	98.4	133.6	120.3	108	95.2	91.6
MAX	466.4	548.1	307.4	280.3	340.1	313.5	286.2	265.2	251.4
Modified Closed-shell systems[b]									
MSE	47.6	182.7	30.5	28.1	68.3	53.1	37.2	20.4	2.9
MAE	48.2	182.7	49.7	36.1	70.3	56	44.3	35.6	41.1
RMS	55.6	199.7	61.3	43.5	79.4	66.6	56.5	43.3	55.7
MAX	122.7	375.5	154.9	100.9	204	189	174	102.4	151.7
Modified Combined Set[a]									
MSE	58.7	197.8	74.2	82.5	67.8	46.8	24.3	-0.7	-31.3
MAE	59.1	199.5	132.2	117.3	69.3	56.1	50.4	53.4	72.6
RMS	68.9	226.2	274.9	265.4	79.2	66.3	65.6	80.1	132.9
MAX	152.8	588.7	1345.7	1239.5	204	189	174	321.6	660.5

[a] For 25 molecules: without $^1CH_2$, HOF, HOOH, N_2H_2.

[b] For 13 molecules: without $^1CH_2$, HOF, HOOH, N_2H_2.

Table 5. Calculated $\langle S^2 \rangle$ values and statistical errors from OD, HF, RIMP2, SOSMP2 and O2 (c_{OS} = 1.2) theories with 6-311G(2df,2pd) basis set for the 12 simple doublet radicals.

	OD	HF	RIMP2	SOSMP2	O2(1.2)
BO	0.7519	0.8039	0.7954	0.7929	0.7530
CF	0.7503	0.7618	0.7563	0.7558	0.7517
CH	0.7502	0.7589	0.7543	0.7539	0.7514
CN	0.7525	1.1560	1.1498	1.1493	0.7523
CO^+	0.7533	1.0034	0.9675	0.9645	0.7552
F_2^+	0.7504	0.7660	0.7571	0.7563	0.7501
FH^+	0.7501	0.7548	0.7522	0.7519	0.7511
N_2^+	0.7509	0.7531	0.7511	0.7509	0.7504
NO	0.7505	0.7947	0.7840	0.7831	0.7508
O_2^+	0.7503	0.7627	0.7554	0.7548	0.7500
OF	0.7506	0.7700	0.7639	0.7633	0.7510
OH	0.7501	0.7559	0.7528	0.7525	0.7511
MAPE (%)[a]	0.1230	6.9672	6.0485	5.9580	0.1876
RMS	0.0014	0.1400	0.1325	0.1319	0.0020
MAX	0.0033	0.4060	0.3998	0.3993	0.0052

[a] MAPE: mean absolute percentage error.

Table 6. Calculated $\langle \hat{S}^2 \rangle$ and total energy (in Hartrees) of the phenalenyl radical and one of the isomers of the phenalenyl σ dimers (RR1) (see Ref. 11) using unrestricted HF, RIMP2 and O2 (c_{OS} = 1.2) theories with 6-31G* basis set. Also shown is the resulting non-counterpoise corrected binding energy (BE) in kcal/mol.

	Method	$\langle \hat{S}^2 \rangle$	Total Energy	BE[a]
Phenalenyl radical[b]	HF	2.0764	-497.583 769	
	RIMP2	1.9238	-499.154 067	
	O2 (1.2)[c]	0.7634	-499.032 361	
Phenalenyl σ dimer (RR1)[b]	HF	3.1991	-995.118 191	30.97
	RIMP2	2.9198	-998.317 634	-5.96
	O2 (1.2)[c]	0.0000	-998.099 082	-21.56

[a] BE = E_{RR1} - 2 $E_{Phenalenyl}$.

[b] Calculation carried out on the unrestricted HF/6-31G* optimized structure obtained from Ref. 11.

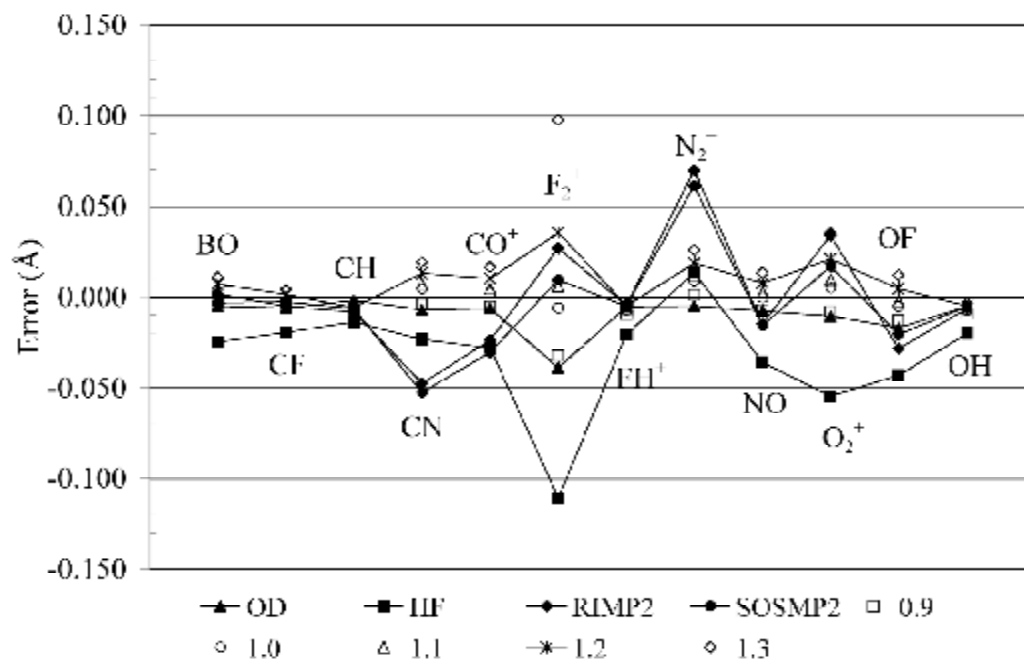[c] Energy evaluated using the unrestricted HF orbitals as the initial reference.

Figure 1. Plot of the deviation of the OD, HF, RIMP2, SOSMP2 and O2 ($c_{OS} = 0.9 - 1.3$) optimized bond lengths from the respective experimental values for the 12 doublet radicals using 6-311G(2df,2pd) basis set.
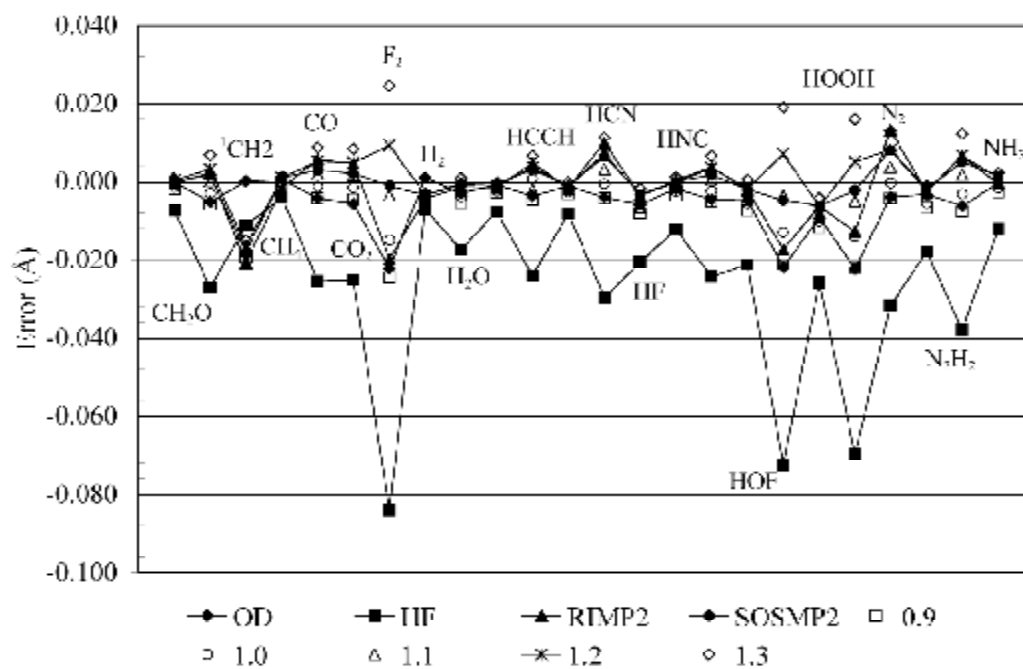
Figure 2. Plot of the deviation of the OD, HF, RIMP2, SOSMP2 and O2 ($c_{OS} = 0.9 - 1.3$) optimized bond lengths from the respective experimental values for the 17 closed-shell molecules using 6-311G(2df,2pd) basis set.
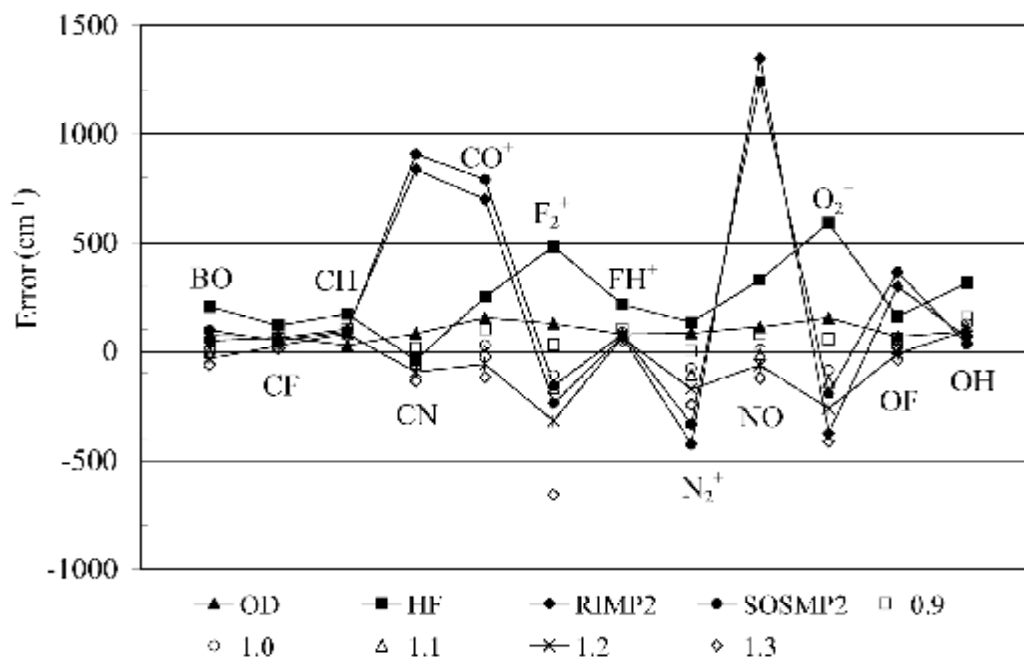
Figure 3. Plot of the deviation of the OD, HF, RIMP2, SOSMP2 and O2 ($c_{OS} = 0.9 - 1.3$) harmonic vibrational frequencies from the respective experimental values for the 12 doublet radicals using 6-311G(2df,2pd) basis set.
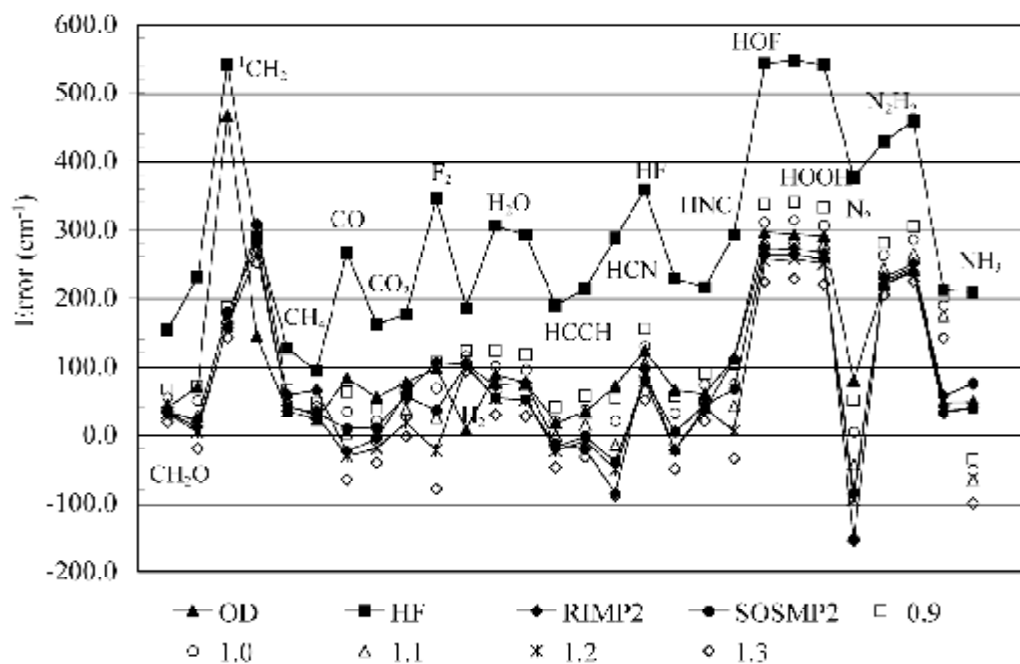
Figure 4. Plot of the deviation of the OD, HF, RI-MP2, SOS-MP2 and O2 ($c_{OS} = 0.9 - 1.3$) vibrational frequencies (only the top two largest errors are shown for all polyatomic molecules) from the respective experimental values for the 17 closed-shell molecules using 6-311G(2df,2pd) basis set.